

Title	動物のGPSデータと方向統計学 (生物流体力学における流れ構造の解析と役割)
Author(s)	島谷, 健一郎
Citation	数理解析研究所講究録 (2014), 1900: 63-67
Issue Date	2014-06
URL	<a href="http://hdl.handle.net/2433/195928">http://hdl.handle.net/2433/195928</a>
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

## 動物の GPS データと方向統計学

統計数理研究所 島谷健一郎

### GPS data of animal movement and directional statistics

The Institute of Statistical Mathematics Ichiro Ken Shimatani

統計モデリングが可能にしてくれることのひとつに、与えられた複数のデータを、まず定性的に分類し、次に各カテゴリーの中で定量的に評価する、というものがある。本稿では、この決して広く知られていないわけではない統計モデリングの役割について、動物の移動軌跡データを題材に紹介する。

#### 1. 軌跡データ

バイオロギング(動物装着型データロガーを用いた動物行動学)や高速カメラなどの技術の発展により、今日、様々な動物の詳細な移動軌跡がデジタルデータとして、蓄積されている。図 1 は GPS ロガーデータの一例で、岩手県三貫島の営巣地から太平洋沖を往復したオオミズナギドリの 1 日の移動軌跡(データは 0.5 秒ごと)である。

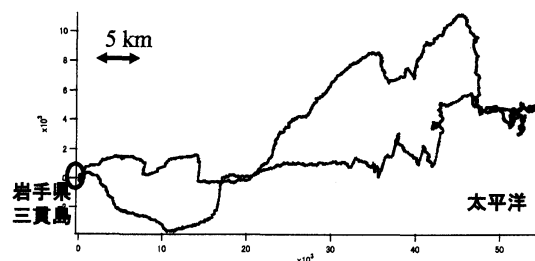


図 1 あるオオミズナギドリの 1 日の移動軌跡

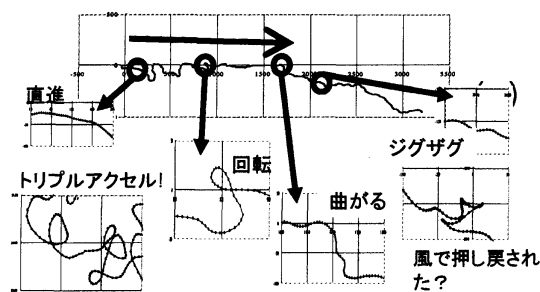


図 2 動物の軌跡はあまりに多様で記述統計もモデリングも難しい

データの一部を拡大してみると、図 2 のように様々なパターンを含んでいる。こうした移動それぞれに固有の目的などがあるはずで、そうした動物の意志を尊重したモデルで動物の行動を説明してみたいと思うのが動物研究者の性である。しかし、このような多様な行動をたった 1 つのモデルで説明するというのは無理がある。そこでまず、軌跡の中で最も多く見られる直線的な部分に注目する(図 3)。いかにもそっちに行きたくて飛んだために残された軌跡に見える。

そこで統計学の手法が必要となるのだが、よく用いられる方法は、各軌跡に対して様々な統計量(始点と終点の距離、始点と終点を結ぶ方向、平均速度、平均角速度、等々)

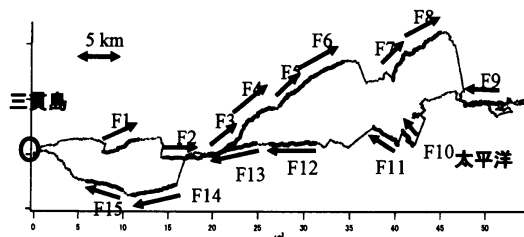


図3 動物軌跡には目標を持った直線的な動きが多く見られる

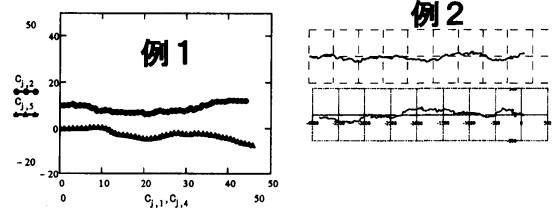


図4 軌跡の特徴を人の目で判断することは難しい。2本の特徴を正確に指摘できる人はいるだろうか。

を計算し、多変量解析を用いた分類である。ただ、こうした手法で分類しても、それはあくまで折れ線という図形の分類であって、鳥の意志は反映されていない。一方、動物の意志を反映したモデルを用いる統計手法では、最尤法でパラメータを最適化し赤池情報量規準による評価でモデルを選択することで、鳥の意志に応じた定性的評価（分類）と、その意志に関する定量的評価が可能となる。もちろん、定性的評価も定量的評価も考えたモデルが正しいと仮定した場合の話であり、ほとんどの場合、人が考えたモデルは動物の意志とほど遠いものでしかないだろう。それでも、モデルのどこがどうデータを説明できないかを精査することでモデルは順次改善され、少しずつ真理に近づいていく。

## 2. Kato の円周自己回帰モデル

動物の目的方向を有する行動のモデルとして、ここでは Kato の円周自己回帰モデル (circular auto regression model, Kato 2010; Shimatani et al. 2012)

$$\theta_t = \alpha + 2 \tan^{-1} \{w \tan((\theta_{t-1} - \alpha)/2)\} + e_t$$

を用いる。ここで、 $\theta_{t-1}$  は時刻  $t-1$  と  $t$  の間の進行方向で、それが次の  $\theta_t$  では、目標方向  $\alpha$  のほうへ修正される (図 5)。  $w$  が 0 に近いと一気に  $\alpha$  に近い方向に修正され、  $w = 0$  では完全に  $\alpha$  へ向かう ( $-1 \leq w \leq 1$ )。但しそれには確率論的エラーが伴い、それが項  $e_t$  である。  $e_t$  は何らかの確率分布に従うと仮定する。  $e_t$  は角度なので、角度 (円周) の

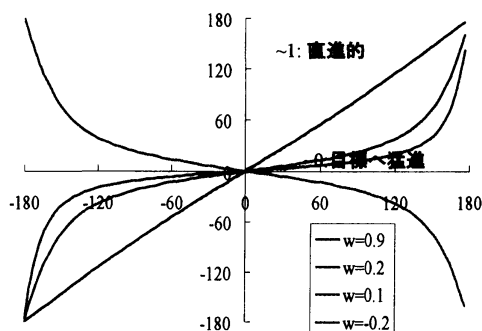


図5 Kato の円周自己回帰モデルの期待値のグラフ。横軸は時刻  $t-1$  の方向、縦軸は時刻  $t$  の方向

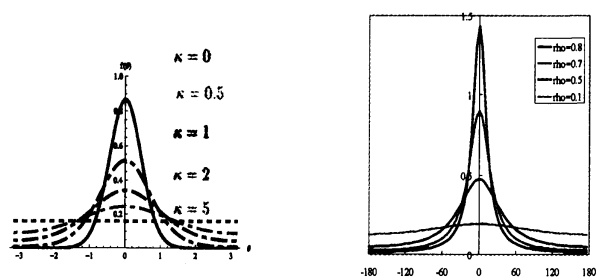


図6 円周上の確率分布の例。左は von Mises 分布、右は wrapped Cauchy 分布

確率分布が望ましく、ここでは von Mises 分布と wrapped Cauchy 分布を用いる (図 6)。それぞれの確率密度関数は以下のように与えられる。

von Mises 分布  $f_{VM}(y; \mu, \kappa) = \exp\{\kappa \cos(y - \mu)\} / 2\pi I_0(\kappa)$

wrapped Cauchy 分布  $f_{WC}(y; \mu, r) = (1 - r^2) / \{2\pi(1 + r^2 - 2r \cos(y - \mu))\}$

確率分布の分散 (厳密には角度の確率分布なので分散に対応する別な統計学の概念が必要。直観的には図 6 に見られる平均方向からの散らばり具合) が大きいとき (von Mises 分布ではパラメータ  $\kappa$  の値が小さいとき、wrapped Cauchy 分布ではパラメータ  $r$  の値が小さいとき)、実際の方向は上記とずれることになる。図 7 にこのモデルで作成した軌跡の例を示す。

$w$  が 1 に近いと、進行方向はほとんど  $\theta_t$  のままで  $\alpha$  方向へは微修正しかされない。図 7 にいろいろなパラメータのときのシミュレーションで作った軌跡の例を示す。

$w = 1$  のときは (1) は

$$\theta_t = \theta_{t-1} + e_t$$

となり、これはランダムウォーク (動物行動学の世界では、これを correlated random walk (CRW) と呼ぶ場合が多い) である。

実際のデータにこのモデルを適用する場合、最尤推定は容易である。尤度式は、データが  $\{\theta_t\}$  ( $t = 1, 2, \dots, n$ ) で与えられたとき、使う確率分布に応じてその密度関数を用いて

$$L_{VM}(\alpha, w, \kappa) = \prod_{t=2}^n f_{VM}(\alpha + 2 \tan^{-1}\{w \tan((\theta_{t-1} - \alpha)/2)\} - \theta_t)$$

$$L_{WC}(\alpha, w, r) = \prod_{t=2}^n f_{WC}(\alpha + 2 \tan^{-1}\{w \tan((\theta_{t-1} - \alpha)/2)\} - \theta_t)$$

で与えられる。この対数が最大となるパラメータ値を何らかの計算ソフトで数値計算すればよい。von Mises 分布では変形ベッセル関数の計算が必要だが、これも今日の計算

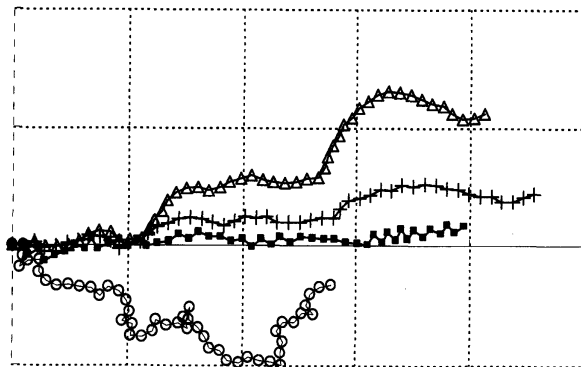


図7 Katoの円周自己回帰モデルで作成した軌跡の例。速さは一定値1とした。 $(w, \kappa)$ の値は+: (0.2, 6), O: (0.2, 2), Δ: (0.7, 6), ■: (-0.7, 6)

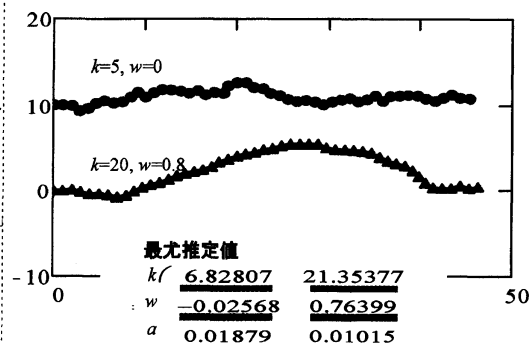


図8 Katoの円周自己回帰モデルで作成した軌跡とパラメータを最尤推定した結果の例。

ソフトの大半が  $\text{=I0()}$  などのコマンドを書くだけで一瞬で求めてくれる。

図 8 の 2 本の軌跡について von Mises 分布のモデルで最尤法を実行すると、図の中に示したような最尤推定値が得られる。これらは元々図内に示した数値で作成した軌跡であり、確かに元のモデルを（ほぼ）正しく推定している。図 7 の中では、下は確率分布の分散は小さいが  $w$  は 1 に近いやや上方にずれていく  $\Delta$  の軌跡、上は分散が大きい + マークの下の方をうろうろしている軌跡に近い。そういわれて図 8 をよく見ると、確かに上はぶるぶる震えた感じを伴うし、下は 1 回 1 回はまっすぐだが 1 度上を向いたらしばらく誤った方向に進んでしまうという点で、図 7 の  $\Delta$  の軌跡に似ている。但し、こうした「感想」は、あくまでモデルと最尤推定値を知ったから言えるのであって、何ものなしに図 8 だけを見て、こうした特徴を言い当てられるものだろうか。

図 9 の 2 本となると、人の目による識別はさらに難しい。実は、上は von Mises 分布、下は wrapped Cauchy 分布を用いたモデルのシミュレーションによる軌跡で、それぞれのパラメータは目的方向（ここでは  $\alpha = 0 =$  東）への進行速度が同じになるように選んだ ( $r = 0.9$  と  $\kappa = 5$  が対応する根拠については Mardia & Japp (1999)などを参照)。図 9 中に最尤推定した結果も示した。パラメータ数が同じなので AIC による相対評価と最大対数尤度による評価は同じなので、図には最大対数尤度値を示してある。それぞれ、正しいモデルで使った確率分布を用いたモデルの方が高い最大対数尤度をとっている。

$w$  は同じでも確率論的エラーは、wrapped Cauchy ではたいていほぼ思惑通りだが稀に大きく外す、von Mises では大体同じように外す。言われてみると確かに軌跡の中で下は大きく外している部分が目立つ。しかしそれもやはり最尤推定をして軌跡を見直して初めて気づくことではなかろうか。

図 10 の 2 本に von Mises 分布を用いたモデルについて最尤法を実行すると、図 10 のような最尤推定値が得られる。真値は上は  $(w, \kappa) = (-0.8, 10)$  下が  $(0, 5)$  である。同じように震えた感じだが、上は動物自らの意志で生じた震え(目標方向と対称な反対方向へ向きを変える)、下は確率論的エラーによるもので、全く異質である。こうした識別も、最尤法で可能となるのである。

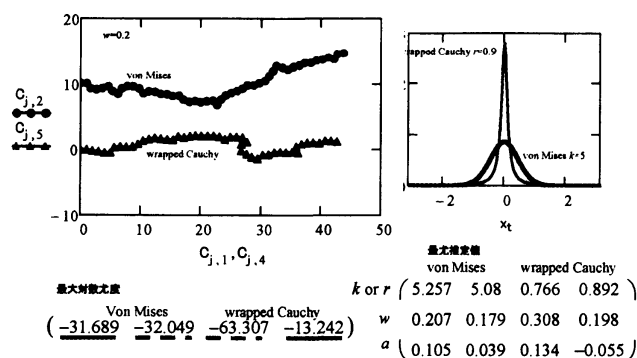


図 9 von Mises と wrapped Cauchy 分布の違い

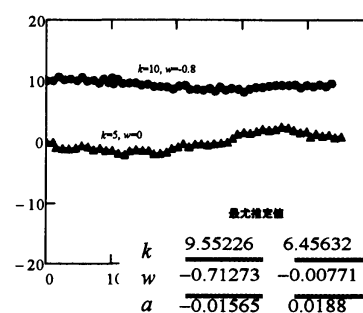


図 10  $w < 0$  の場合

$w = 1$  のときは correlated random walk に帰着するが、この場合、パラメータは ( $\alpha$  も不要になるので) 2 つ減る。そこで、与えられた軌跡データに対し、von Mises 及び wrapped Cauchy 分布を用いた円周自己回帰モデルと、correlated random walk を適用し最尤法を実行する。パラメータ数が異なるので、赤池情報量規準(AIC)  $AIC = -2$  (最大対数尤度)  $+ 2$  (パラメータ数) を計算し、AIC 値が最小のモデルを選択することで、まず、軌跡を定性的に 4 つに分類し、各カテゴリーの中で、最尤推定値を用いて定量的に評価することが可能となる。

図 11 は上記の 4 つのモデルを用いて作成した 8 本の軌跡だが、最尤法と AIC は、すべて真のモデルを選択し、最尤推定値も真値に近いものになっている。

こうして、1 つのモデルにより、様々な軌跡を特徴に応じて定性的に分類し、さらに定量的に評価する枠組みが確立された。次のステップは、この基本的モデルではたいていの実際の行動軌跡は説明できないはずなので、どこがどうモデル予測から外れているかを検定する枠組みの確立と、モデルの拡張である。

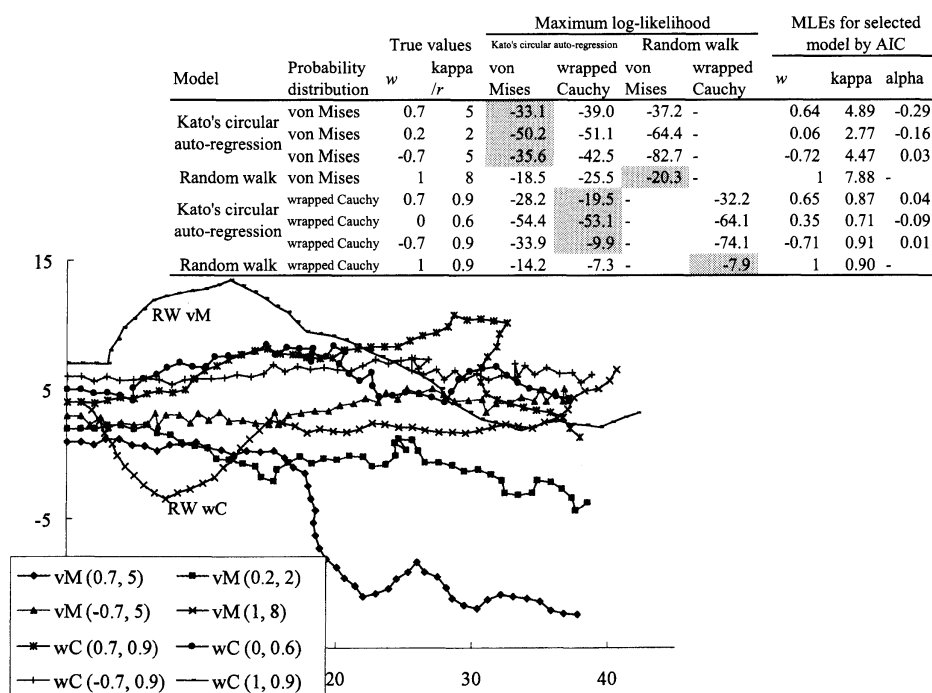


図 11 4 つのモデルから様々なパラメータ値で描いた軌跡の例と、その最尤推定値及び AIC 値(右上の表)。

#### 参考文献

Kato S (2010) A Markov process for circular data. J Roy Stat Soc B 72: 655–672.

Mardia KV, Jupp PE (1999) Directional Statistics. Chichester: Wiley.

Shimatani Ken Ichiro, Yoda, K., Katsumata, N., Sato, K. (2012) Toward the Quantification of a Conceptual Framework for Movement Ecology Using Circular Statistical Modeling, PLoS One, doi:10.1371/journal.pone.0050309.